

Using Geostatistical Methods in the Analysis of Public Health Data: The Final Frontier?

Linda J. Young and Carol A. Gotway

Abstract Geostatistical methods have been demonstrated to be very powerful analytical tools in a variety of disciplines, most notably in mining, agriculture, meteorology, hydrology, geology and environmental science. Unfortunately, their use in public health, medical geography, and spatial epidemiology has languished in favor of Bayesian methods or the analytical methods developed in geography and promoted via geographic information systems. In this presentation, we provide our views concerning the use of geostatistical methods for analyzing spatial public health data. We revisit the geostatistical paradigm in light of traditional analytical examples from public health. We discuss the challenges that need to be faced in applying geostatistical methods to the analysis of public health data as well as the opportunities for increasing the use of geostatistical methods in public health applications.

1 Introduction

Analysis of spatial data has come to be important for many studies in public health, medical geography and spatial epidemiology. Whereas geostatistical methods have been used extensively in a variety of disciplines, including mining, agriculture, meteorology, hydrology, geology and environmental science, they have found only limited application in health studies where Bayesian methods and analytical methods developed in geography and implemented in geographic information systems have dominated. Here, we consider some of the challenges encountered in our efforts to use geostatistical methods for analyzing spatial public health data and some of the solutions that have been proposed. This is not meant to be a comprehensive list, but one that reflects our experiences and identifies needs for additional research.

L.J. Young (✉)

Department of Statistics, 404 McCarty Hall C, P.O. Box 110339, University of Florida, Gainesville, FL 32611-0339, USA

e-mail: LJYoung@ufl.edu

C.A. Gotway

Centers for Disease Control and Prevention, Office of Workforce and Career Development, 1600 Clifton Rd, NE, MS E-94, Atlanta, GA 30333, USA

e-mail: cdg7@cdc.gov

2 Motivating Study

Our work with Florida's Environmental Public Health Tracking (EPHT) effort provides the motivating study (Young et al., 2008). Part of Florida's efforts to move toward implementation of EPHT is to develop models of the spatial and temporal association between myocardial infarctions (MIs) and the changing levels of ozone in outdoor air for Florida. To accomplish this, as with the majority of studies relating environmental changes to public health, especially those that are national or regional in scope, the analysis is based on pre-existing data. Florida's Department of Environmental Protection (FDEP) provided ozone measurements, recorded from a network of 48 air monitors placed throughout the state. Florida's Agency for Health Care Administration (AHCA), consistent with a data sharing agreement, provided all admissions to Florida's public and private hospitals where either the primary or secondary cause of admission was MI (*International Classification of Diseases*, 10th Revision (ICD-10) codes 410.0–414.0 [World Health Organization]). ACHA also provided both the zip code and county of residence for each patient's record and selected patient demographic information, including sex, age, and race/ethnicity. Selected sociodemographic data (age, race/ethnicity, sex, education) were obtained from the U.S. Census Bureau. Additional sociodemographic data were obtained from CDC's Behavioral Risk Factor Surveillance System (BRFSS). For March, 2001, the number of MI admissions per 10,000 population and the 48 ozone monitors functioning that month, are displayed in Fig. 1 (see Young et al., 2008 for full details).

3 Challenges for Public Health

3.1 Spatial Support

As illustrated in our Florida study, increasingly interest extends beyond the simple reporting of incidence or risk and turns to relating these responses to potential explanatory variables. As is also common, the variables used in each of these studies were collected from disparate sources and must be linked on a common set of spatial units for analysis. Moving from one set of spatial units to another can result in several challenging change of support problems (see Gotway and Young 2002 for a review). Most of the early geostatistical work on change of support problems was motivated by mining applications in which the inferential unit of interest was a block of ore. The rectangular shape of blocks made it possible to use a regular grid to discretize the blocks into points and approximate the integrals needed for block kriging using just a relatively few number of points. However, applications in the public health field call for a reassessment and extension to this and other geostatistical approaches.

First, the "blocks" are seldom rectangular in shape or consistent in size. As an example, note that the Florida counties (Fig. 1) vary considerably in size and are

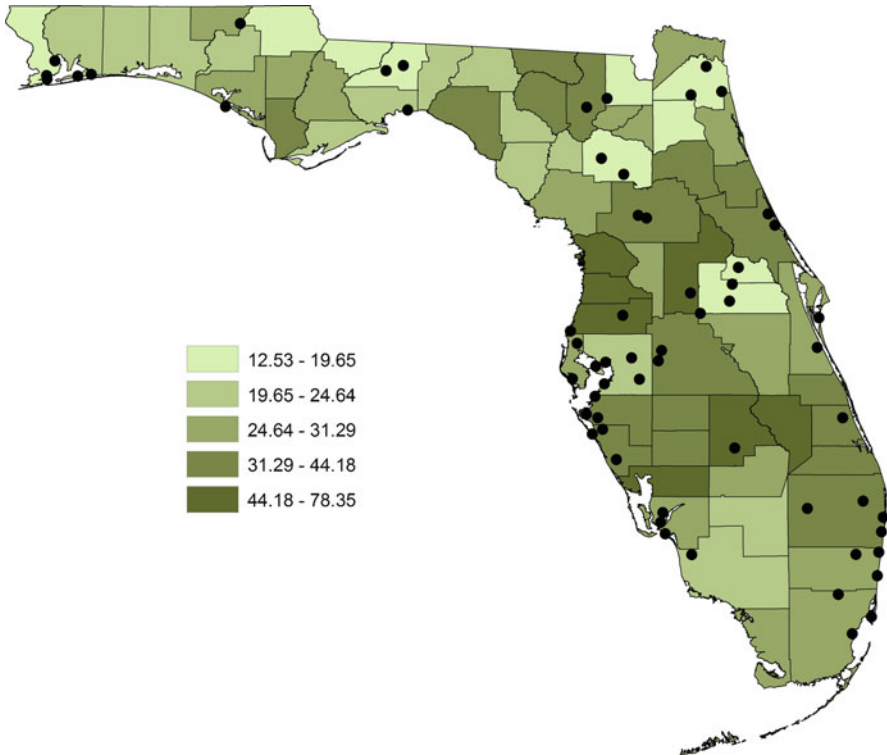


Fig. 1 The number of MI cases per 10,000 population recorded for each Florida county during March 2001 and the location of the ozone monitors functioning during that month

irregular in shape. This is typical of the postal codes, political boundaries, and census administration units often used in public health studies. Here we wanted to use the 48 ozone monitors functioning during March, 2001, to obtain an average maximum ozone value for each county. A regular grid was placed across all of Florida, and three counties did not have any grid points or monitors falling within them. One option was to make the grid very fine. This would have slowed computations tremendously and is inefficient because the larger counties would be over-characterized. Alternatively, we augmented the grid with a finer sub-grid for those three counties. Is this the best approach?

A second challenge results from the different change of support problems encountered in public health, often arising from data confidentiality concerns. Typically, the change of support problem is not one of upscaling (or aggregation). As an example, the incidence of low birth weight babies is available at the county level, but interest lies in incidence of low birth weight babies at the census tract level (Gotway and Young, 2007). This downscaling (or disaggregation) should ideally preserve the pycnophylactic property (Tobler, 1979) that the number of low birth weight babies from the census tracts within a county should equal the county total.

As another example, health outcomes are generally reported on the zip code level, but demographic data are provided on the census tract level. Here the spatial units from the two sources overlap and “side-scaling” is needed to properly assign demographic data to zip code units. [Gotway and Young \(2007\)](#) generalized geostatistical methods, which have historically focused on upscaling, for use in general change-of-support problems, including upscaling, down scaling, side-scaling, and intensity estimation.

We should acknowledge that, in many studies, the change-of-support problem is ignored, primarily due to the complexity of the solution proposed and the lack of software. For example, when working with the March ozone data, one suggested approach was to identify proxy monitors that would represent the ozone values for any county without a monitor, an approach that does not address support, but greatly simplifies the computational issues. Similarly, the non-geostatistical methods that have been proposed for change-of-support often do not consider the support of the data (e.g., proportional allocation, centroid smoothing). Further, instead of explicitly accounting for support in a geostatistical approach, [Diggle and Robeiro \(2007\)](#) suggest that an alternative approach is to partition the spatial region into n discrete spatial units, each with a response variable y_i , $i = 1, \dots, n$, and then model the multivariate distribution for the random variable Y_i . Undoubtedly, accounting for support in spatial analysis is challenging, both theoretically and computationally. However, in mining, accounting for support was found to be critically important and predicting a spatial average is very different from simply predicting an average at a point. The lesson likely holds for public health as well, and we should learn from the mining experience where accurate block-grade predictions and inferences are critical to the profits of the industry.

3.2 *Discrete Distributions*

Geostatisticians working in public health and other application areas have responded to the need for new methods for discrete distributions, especially the Poisson and binomial distributions. Unlike the Gaussian distribution, the variance of any discrete probability distribution depends on the mean. For the Poisson, the mean and variance are equal; for the binomial, the variance is equal to the mean multiplied by a constant that is less than one. The Box-Cox family of transformations includes transformations which stabilize variance, and using the appropriate transformation from this family in a trans-Gaussian kriging ([Schabenberger and Gotway, 2005](#), pp. 270–277) formulation may work well. However, models that explicitly account for the variance–mean relationships inherent in many discrete distributions are warranted for applications to other disciplines such as public health.

Poisson kriging was developed by [Monestiez et al. \(2005, 2006\)](#) for mapping the spatial distribution of fin whales and used to predict cancer mortality rates in a public health setting by [Goovaerts \(2005\)](#). In the public health context, we have $Z(B_i)$, the count or total number of disease cases over the i th region with population $n(B_i)$

at risk. Thus, $R(B_i) = Z(B_i)/n(B_i)$ is the incidence proportion for region B_i . Assume that $\{\lambda(\mathbf{s})|\mathbf{s} \in D \subset \mathfrak{R}^2\}$ is an unobserved intensity process, with $\lambda(\mathbf{s}) \geq 0$ for all \mathbf{s} in D . Assume this process has mean μ_λ and covariance function $C_{\lambda\lambda}(\mathbf{s}_i, \mathbf{s}_j)$. Further assume that, conditional on this process, the observed frequencies (counts), $Z(B_i)$, associated with an areal region B_i are independent Poisson random variables with means and variances both equal to $\lambda(\mathbf{s})n(B_i)$. If we assume a linear prediction function for $\lambda(\mathbf{s})$, then the predictor of the intensity process at location \mathbf{s}_0 is

$$\hat{\lambda}(\mathbf{s}_0) = \sum_{i=1}^N w_i R(B_i),$$

where N is the number of regions and optimal weights, w_i , can be obtained by solving

$$\sum_{k=1}^N w_k \left[C_{RR}(B_i, B_k) + \delta_{ik} \frac{\mu^*}{n(B_i)} \right] + m = C_{\lambda R}(\mathbf{s}_0, B_i), \quad i = 1, \dots, n$$

$$\sum_{k=1}^N w_k = 1.$$

Here $\delta_{ik} = 1$ if $B_i = B_k$ and 0 otherwise, μ^* is an estimate of the mean of $R(\cdot)$, and m is a Lagrange multiplier. A key to the estimation process is estimation of the point-support covariance function from which the cross-covariance function between the intensity process and the observed frequencies is determined. In an effort to adjust for heterogeneous variances, Monestiez et al. (2005, 2006) proposed weighting the difference pair by the corresponding population sizes. Extending the ideas of Mockus (1998), Goovaerts (2008) proposes an iterative deconvolution method. Here too is a change of support problem: $\lambda(\mathbf{s}_i)$ is assumed to be of point-support, but $R(B_i)$ is aggregated over areal regions. Binomial kriging (McNeill, 1991) has a similar derivation and leads to comparable challenges.

Gotway and Stroup (1997) developed models for generalized linear models, of which the Poisson and binomial are special cases. In an approach similar to that of trans-Gaussian kriging, they used Taylor series to linearize the problem so that the usual kriging predictor is optimal, but with variance-mean relationships built into models for spatial dependence. Gotway and Wolfinger (2003) compare these models to those conditioned on a latent process as in Poisson kriging, binomial kriging, and model-based geostatistics. Their results indicate that while conditionally-specified models can be used to build complicated, non-stationary models, they tended to under-predict both counts and rates and may severely over-estimate prediction uncertainty for data sets with moderate-to-large marginal spatial autocorrelation. The marginal models allow us to move away from any Gaussian assumptions and employ methods similar in form to least squares estimation. However, the estimation algorithm was not as stable for these models, and the predictions tended to vary more than those from the conditional model. Ordinary or universal kriging, with a semivariogram weighted inversely proportional to the assumed variance of the data

(in this case, inversely proportional to $n(B_i)$) worked surprisingly well, demonstrating what most geostatistical practitioners have observed time and again: ordinary kriging is relatively robust to a variety of violations in assumptions. Although predictions may not be theoretically optimal, they are not grossly inaccurate either. Nevertheless, models that better describe the nature of the problem and the properties of the data are intuitively more appealing.

With both Poisson and binomial kriging, and marginal generalized linear models, two issues have yet to be fully addressed. One important issue is that, in geostatistical modelling, we are working with multivariate data and we need an underlying joint multivariate distribution for valid inference. Although this may appear to be a simple theoretical nuisance, the lack of such a multivariate distribution can cause difficulties, such as “covariance” matrices that are not positive definite, numerical instability, and order-relations problems, in some practical applications. Herein lies the problem with the non-parametric indicator approaches and Poisson, binomial, and generalized linear model approaches. A classic example is indicator kriging which predicts probabilities, which, theoretically, should be contained in $[0,1]$. However, any user of indicator methods has obtained predicted probabilities outside this range.

A number of the challenges arise in constructing non-Gaussian, multivariate distributions with specified correlation structure, marginal distributions, and conditional distributions (see [Schabenberger and Gotway, 2004](#), pp. 192–195, for a full discussion). Constraints on the correlation exist for many multivariate distributions that are not constructed from an underlying multivariate Gaussian distribution. As an example, the multivariate binomial permits only negative correlations ([Mardia 1970](#)). For other models, no such multivariate distribution exists. For example, no multivariate distribution exists having both marginal and conditional distributions of Poisson form ([Mardia, 1970](#)).

Generating multivariate distributions sequentially from specified conditions overcomes some of these difficulties. In Bayesian hierarchical modeling, this sequential conditioning approach is used to generate fairly complex multivariate distributions, but the properties of the resulting distribution may not always be clear. As an example, suppose $Z_1(\mathbf{s})$ is a second-order stationary process with $E[Z_1(\mathbf{s})] = 1$ and $\text{Cov}[Z_1(\mathbf{u}), Z_1(\mathbf{u} + \mathbf{h})] = \sigma^2 \rho_1(\mathbf{h})$. A simplified version of a common model used for modeling and inference with count data is obtained by conditioning $Z_2(s)$, a white noise process with mean and variance given by

$$E[Z_2(\mathbf{s})|Z_1(\mathbf{s})] = \exp\{\mathbf{x}(\mathbf{s})'\boldsymbol{\beta}\}Z_1(\mathbf{s}) \equiv \mu(\mathbf{s}), \quad \text{Var}[Z_2(\mathbf{s})|Z_1(\mathbf{s})] = \mu(\mathbf{s})$$

on $Z_1(\mathbf{s})$. The marginal mean $E[Z_2(\mathbf{s})] = \exp\{\mathbf{x}(\mathbf{s})'\boldsymbol{\beta}\}$, depends only on the unknown parameter $\boldsymbol{\beta}$, and the marginal variance, $\text{Var}[Z_2(\mathbf{s})] = \mu(\mathbf{s}) + \sigma^2\mu(\mathbf{s})^2$, allows overdispersion in the data $Z_2(\mathbf{s})$, making the model attractive. Now, consider the marginal correlation of $Z_2(\mathbf{s})$

$$\text{Corr}[Z_2(\mathbf{s}), Z_2(\mathbf{s}+\mathbf{h})] = \frac{\rho_1(\mathbf{h})}{\left[\left(1 + \frac{1}{\sigma^2\mu(\mathbf{s})}\right)\left(1 + \frac{1}{\sigma^2\mu(\mathbf{s}+\mathbf{h})}\right)\right]^{1/2}}$$

If σ^2 , $\mu(\mathbf{s})$, and $\mu(\mathbf{s} + \mathbf{h})$ are small, $\text{Corr}[Z_2(\mathbf{s}), Z_2(\mathbf{s} + \mathbf{h})] \ll \rho_1(\mathbf{h})$. Thus, while the conditioning induces both overdispersion and autocorrelation in the Z_2 process, the marginal correlation has a definite upper bound and so may not be a good model for highly correlated data. Most Bayesian models have a similar constraint built in, although it is often difficult to test either theoretically or empirically.

The second fundamental issue is that the marginal variance and the covariance function depends on $n(\mathbf{B}_i)$ (e.g., Goovaerts, 2005; Monestiez et al., 2005, 2006). Thus, neither Poisson nor binomial kriging are based on an intrinsically stationary process. Weighting the empirical semivariogram by factors that are inversely proportional to the standard deviation of the data (Goovaerts, 2005; Monestiez et al., 2005, 2006) ameliorates the problem. However, the semivariogram of the data process is only estimable (and arguably only defined) for intrinsically stationary processes. This problem of non-stationarity affects the validity of all the geostatistical tools such as measures of autocorrelation, spatial prediction, and geostatistical simulation methods. Moreover, covariates may not be spatially continuous and are often categorical. Thus, non-stationarity arises in two ways: differing populations and the need to adjust for covariates. Although most geostatistical tools are robust to departures from the assumption of stationarity, the lack of a more general paradigm may prevent their wide-spread adoption in public health.

More sophisticated models for prediction with discrete distribution have also been developed, including disjunctive kriging methods and isofactorial models (e.g., Rivoirard, 1994) and Bayesian methods (Diggle et al., 1998). Unfortunately, none of these approaches is ready for routine use, and the general Bayesian methods have yet to be extended to complex change of support problems.

Given the above discussion, the reasons for the popularity of the multivariate Gaussian distribution are evident. It has a closed form expression, permits pairwise correlations in $(-1, 1)$, each (Z_i, Z_j) has a bivariate Gaussian distribution, all marginal distributions are Gaussian, and all conditional distributions are Gaussian. Moreover, tractable multivariate distributions, such as the multivariate lognormal and the multivariate t -distribution can be derived from the multivariate Gaussian. The Gaussian distribution has truly earned its unique place in geostatistical theory. Thus, for our motivating study, instead of using methods developed for discrete distributions, the incidence of MI at the county level was indirectly standardized by age, sex, and education to the Florida population and the standardized event ratio (MI SER) computed. The MI SER was log-transformed (denoted by $\ln(\text{SER})$ because the natural logarithm was taken) so that the assumptions of linear regression (normality and constant variance) would be more nearly met.

3.3 Spatial Regression

The traditional analytical approach, referred to here as *global regression*, is to conduct a multivariate linear regression analysis relating the health outcome to potential predictors with adjustments for sociodemographic variables (e.g., education,

income, and percentage of smokers). For our study, a weighted regression was conducted with the weight being equal to the expected MI SER, and the coefficient on ozone was exponentiated to obtain the relative MI SER from the regression.

Just as ozone levels and the number of MI cases can vary over the state, the relative MI SER could also vary over the state. [Hastie and Tibshirani \(1993\)](#) introduced varying coefficient models, a class of regression and generalized regression functions in which the coefficients are allowed to vary as smooth functions of other variables. [Müller \(1998\)](#) adapted this idea to the spatial case and referred to the approach as *local regression*. Independently, [Brunsdon et al. \(1996\)](#) adapted the idea of varying coefficient models to the spatial case and called their method *geographically weighted regression*. More generally, when regression coefficients are assumed to vary smoothly over space, the models are referred to as spatially varying coefficient models ([Gelfand et al., 2003](#)).

To fit a local regression model, ideas from local smoothing and kernel regression are used to define spatial neighborhoods. The regression is performed by using only data in the spatial neighborhoods. As a consequence, the error terms are not necessarily constant for all locations. Further, because the spatial neighborhoods associated with different points in space overlap, the same data are used more than once to estimate all the spatial regression parameters. Local regression models are appealing because we expect risk to change over space as well as with time, and this can be an important outcome for public health studies. Yet this method has open questions. Because the same data are used more than once to estimate all the spatial regression parameters (β s), a correlation structure is induced among the β s. One consequence of this correlation might be overly smoothed predictions. In our motivating study, the estimated relative MI SERs are much smoother than either the MI SERs or the predicted ozone values. This phenomenon can be observed for other, similar local regression models for both frequentist (as presented here, see also [Nakaya et al., 2005](#)) and Bayesian analyses (e.g., [Waller et al., 2007](#)). As is often the case with Bayesian analyses, the local regression models are overparameterized, and assumptions (e.g., the form of the prior distributions) allow one to proceed with the analyses. In local regression, as in other analyses using overparameterized models, the impact of the assumptions is not fully evident.

Health outcomes are likely to depend on more than one environmental factor (e.g., the ozone levels considered here). This leads us to include other explanatory variables (e.g., PM2.5) in the models. [Wheeler and Tiefelsdorf \(2005\)](#) concluded that, for local regression, multicollinearity among the coefficients at a single location and the overall correlation between coefficients associated with two different explanatory variables (e.g., ozone and PM2.5) can make interpretation of the model coefficients problematic. Their results indicate that the collinearity among local regression coefficients might be present even if the process generating the explanatory variables leads them to be uncorrelated. This collinearity is likely caused by implicit conditions that are placed on the parameters during the estimation process. This is an open question worthy of further research, as is the more general concern of valid inference from all local regression models, because they were designed as exploratory smoothing methods and not inferential statistical tools.

4 Conclusions

Throughout this work, we have been critical of the existing methods as they related to public health studies. Our goal has been to emphasize the vast opportunities for research on important geostatistical issues. Here we want to take time to applaud the authors whose work we have critiqued. Although we have pointed out areas that need further development, we are encouraged that efforts are being made to address complex issues that arise.

Discrete and, more generally, non-Gaussian data are common in public health studies. Satisfactory multivariate non-Gaussian models have severe limitations. Either we do not get the marginal or conditional distributions that are desired or the choice of covariance structures is severely limited. Is the best solution to transform the data so that it is at least approximately normal and to then rely on the robustness of the standard geostatistical methods? Or, even with the disadvantages outlined here, is it better to use methods such as Poisson kriging? Is there a better approach? These are examples of the basic guidance that those working in public health need if geostatistical methods are to find broader application.

Acknowledgements The senior author was partially supported by the Florida Department of Health, Division of Environmental Health and Grant/Cooperative Agreement Number 5 U38 EH000177-02 from the Centers for Disease Control and Prevention (CDC). The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

References

- Brunsdon CF, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal* 28:281–298
- Centers for Disease Control and Prevention (CDC) (2006) Health risks in the United States: behavioral risk factor surveillance system 2006. US Department of Health and Human Services, CDC, Atlanta, GA
- Diggle PJ, Røbeiro PJ (2007) *Model-based geostatistics*. Springer, New York
- Diggle PJ, Tawn JA, Moyeed RA (1998) Model based geostatistics. *Appl Stat* 47:299–350
- Gelfand AE, Kim H-J, Sirmans CF, Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. *J Am Stat Assoc* 98:387–396
- Goovaerts P (2005) Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. *Int J Health Geogr* 4:31
- Goovaerts P (2008) Geostatistical analysis of health data: state-of-the-art and perspective. In: Soares A, Pereira MJ, Dimitrakopoulos R (eds) *geoENV VI –geostatistics for environmental applications*. Springer, The Netherlands, pp 3–22.
- Gotway CA, Stroup WW (1997) A generalized linear model approach to spatial data analysis and prediction. *J Agric, Biol, Environ Stat* 2:157–178
- Gotway CA, Wolfinger RD (2003) Spatial prediction of counts and rates. *Stat Med* 22:1415–1432
- Gotway CA, Young LJ (2002) Combining incompatible spatial data. *J Am Stat Assoc* 97:632–648
- Gotway CA, Young LJ (2007) A geostatistical approach to linking spatially-aggregated data from different sources. *J Comput Graph Stat* 16:115–135
- Hastie TJ, Tibshirani RJ (1993) Varying-coefficient models. *J R Stat Soc B* 55:757–796

- Mardia KV (1970) Families of bivariate distributions. Hafner, Darienn, CT
- McNeill L (1991) Interpolation and smoothing of binomial data for the Southern African Bird Atlas Project. *S Afr J Stat* 25:129–136
- Mockus A (1998) Estimating dependencies from spatial averages. *J Comput Graph Stat* 7:501–513
- Monestiez P, Dubroca L, Bonnin E, Durbec JP, Guinet C (2005) Comparison of model based geostatistical methods in ecology: application to fin whale spatial distribution in northwestern Mediterranean Sea. In: Leuangthong O, Deutsch CV (eds) *Geostatistics Banf 2005*, vol. 2. Kluwer, Dordrecht, The Netherlands, pp 777–786
- Monestiez P, Dubroca L, Bonnin E, Durbec J-P, Guinet C (2006) Geostatistical modeling of spatial distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observations efforts. *Ecol Model* 193:615–628
- Müller WG (1998) Fundamentals of spatial statistics. In: *Collecting spatial data: optimum design of experiments for random fields*. Physica-Verlag, Heidelberg
- Nakaya T, Fotheringham AS, Brunson C, Charlton M (2005) Geographically weighted Poisson regression for disease association mapping. *Stat Med* 24:2695–2717
- Rivoirard J (1994) *Introduction to disjunctive kriging and non-linear geostatistics*. Clarendon, Oxford
- Schabenberger O, Gotway CA (2005) *Statistical methods for spatial data analysis*. CRC Press, Boca Raton, FL
- Tobler W (1979) Smooth pycnophylactic interpolation for geographical regions (with discussion). *J Am Stat Assoc* 74:519–536
- U.S. Census Bureau (2001) *Age: 2000. Economics and statistics administration*. U.S. Department of Commerce, Washington, DC
- Waller LA, Zhu L, Gotway CA, Gorman D, Gruenewald P (2007) Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models. *Stoch Environ Res Risk Assess* 21:573–588
- Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *J Geogr Syst* 7:161–187
- World Health Organization (2005) *International classification of diseases and related health problems (ICD-10)*, 2nd edn. WHO
- Young LJ, Gotway CA, Yang J, Kearney G, DuClos C (2008) Assessing the association between environmental impacts and health outcomes: a case study from Florida. *Stat Med* (in Press). doi: 10.1002/sim.3249